State of Illinois
# ENVIRONMENTAL PROTECTION AGENCY

Mary A. Gade, Director
2200 Churchill Road, Springfield, IL 62794-9276

217/782-6760

February 24, 1995

Mr Brad Bradley
Remedial Project Manager
Office of Superfund (HSRL-6J)
U.S. EPA Region 5
77 West Jackson Boulevard
Chicago, Illinois 60604

**Re: NL/Taracorp Waste Pile Investigation**
**L1190400007**

Dear Brad:

I received the draft report for the waste pile investigation for the NL/Taracorp Site submitted by Woodward-Clyde on January 17, 1995 and have completed my review of the contents. My only question was about the acceptable BTU value for the secondary smelting facility. I called David Pate of Woodward-Clyde and he explained why the pile contents must have a total BTU value less than 5000 in order for it to be accepted by the secondary smelting facility. The report is very informative and will be beneficial to future work on the pile.

If you have any questions, please contact me at 217/785-8729 or at the above address.

Sincerely,

Robert Rogers

Robert Rogers, E.I.T.
Project Manager, NPL Unit
Remedial Project Management Section
Bureau of Land

cc: Eugene Liu, USACOE
David Pate, Woodward-Clyde

# RESPONSES TO COMMENTS

## Memo from Maurice LeVois to Tom Long, July 21, 1994

**Comment ML1, p. 1, Sec. 1.2**: "The term 'control group' implies that there is a clear definition of what, how, and why we are controlling by design or analysis strategy. In this case Pontoon Beach was different with respect to SES and living conditions ... They were not comparable to residents in *our main study area (composed of old houses situated near the proposed cleanup area.* **[emphasis added]** Residents from neighboring areas of Granite City were far more comparable to our target group, ...

"Use of a 'control group' is actually an error in the design of studies of the effects of residential lead, ubless it can be shown that the control group is like the study group in every respect except soil lead level. Our sample of subjects drawn from a more homogeneous population spread over a distinct gradient of soil lead levels is the only sensible study design under these conditions."

**Response to comment ML1:**

Residents from the target area are demographically different from residents in outlying parts of Granite City and from residents in similar housing in Madison, in Venice Township, and in unincorporated parts of Madison County. Differences include socioeconomic status (denoted SES) and race or ethnicity, which have been shown in many other studies (e.g. Stark et al. 1982) to be modifying factors in childhood lead exposure. Unfortunately, data provided by IDPH to EPA do not allow evaluation of the characteristics of the target area with those of other parts of the Madison County study area, since there is no geographical location information in the data base that would allow grouping of subjects in other neighborhoods or communities in Madison County as candidate control groups. The data provided by IDPH to EPA does not even include the community withinn which the subjects are located. The data set includes a crude index of distance from the NL site, but there are parts of Granite City, Venice, or Madison in each ring of distance, with prior observation of these communities suggesting large differences that may affect lead exposure. Therefore, the claim that the other parts of Madison County are appropriate control areas for the target area has not been justified. In view of the potential confounding issues described below, this is a serious deficiency in the design of the study.

We are particularly concerned that little effort appears to have been made by IEHR or by IDPH to identify potential control areas outside of the study area, apart from Pontoon Beach. The characteristics for such an area, which were readily identifiable from many prior studies, would include a gradient of factors with respect to a centrally-located industrial facility such as age of housing, levels and condition of lead-based paint on housing, building condition, race or ethnicity, building condition, income and level of education, but lacking an industrial lead source. A number of small- and medium-sized communities in the Midwest could have been evaluated.

While some care may be needed in matching control communities to target communities, this is a standard method in epidemiologic health studies and the conditions for adequate matching of control groups and exposed groups are well known even to beginning students of epidemiology (see basic texts, e.g. Rothman, 1986). Even though some environmental lead health studies have not used matching control groups, they sometimes have included a post-study analysis that demonstrates the absence of significant confounding with certain identifiable factors measured in the study. The lack of an external control group is sometimes necessary, but to deliberately fail to include an external control group in a lead health study is to make a virtue of necessity, whereas it must be regarded as a design deficiency with potentially serious consequences in interpreting analyses of the data. IDPH has not yet provided EPA with all of data that are necessary to allow a post-study assessment of the Madison County study to determine whether this design deficiency is merely an annoying complication that can be overcome by suitable re-analyses of the data, or a fatal deficency that from the very beginning precluded valid inference from the study.

**Comment ML2, pp. 1-2, Sec. 1.3:** "Re-sampling of blood lead, combined with counselling intervention, resulted in a greater drop of blood lead than expected."

**Response to Comment ML2:**

This is an interesting observation, but it has little basis as a generally valid scientific conclusion, since the observed reduction could have occurred for any number of reasons not related to the intervention. Blood lead decreases may have also occurred in children with similar blood lead concentrations in other households, had the investigators chosen to include control groups in the follow-up study. These investigators seem indifferent to the importance of control groups in establishing valid scientific conclusions, as noted in Response ML1. The quantitative effects of counselling and intervention on child blood lead, and the persistence or lack of persistence of such effects over seasons and years are important questions. The expenditure of resources to obtain data that were guaranteed to be inconclusive by the design (or lack of design) of the followup study is most regrettable.

**Comment ML3, p. 2, Sec. 2.3, para. 2:** "Ten soil samples were collected from the primary play areas in the yard around each house. No soil samples were taken from within the drip'line of the house. A composite soil sample was made from the ten samples. This procedure should have yielded a representative soil sample from the yards and play areas. Since the great majority of the yards were very small, it is highly unlikely that the soil sampling protocol could have yielded unrepresentative soil sample results."

**Response to Comment ML3:**

The absence of drip line samples implies that these measurements are not exactly comparable to those used in calibrating the IEUBK model. Sampling protocols used by EPA and other investigators requirethat some samples be collected within 0.5 to 1 meter of the housing unit outside wall. The dripline or house perimeter samples usually have higher soil lead concentrations than those taken further out in the yard, since they include some rooftop runoff with airborne particles deposited there, along with some of the flaking and chaking exterior lead-based paint when it is present. Dripline soil samples tend to give better prediction of dust lead levels at the house entry areas and interior, so are useful for understanding lead pathways. Since some children play near the house, soil lead samples may also be predictive of child blood lead (Wesolowski et al., 1983). We would therefore expect that lead concentrations in the Madison County soil samples are: (i) less affected by exterior lead-based paint than perimeter soil samples or composite samples containing perimeter soil than in other studies; (ii) less predictive of interior dust lead than soil samples collected in other studies. Our reanalyses of the data suggest that this may indeed have occurred.

**Comment ML4, Sec. 3.1, p.2:** "Employing a nonlinear age covariate in blood lead regression models could increase slightly the amount of blood lead variance accounted for by age. That would have the effect of reducing slightly the amount of variance in blood lead remainng for other variables, such as soil and dust, to explain."

**Response to Comment ML4:**

The purpose of the analytical modelling should have been directed towards the development of models that could be used to better assess the effects of changes in lead exposure on child blood lead. It has long been known that the child's age can be used as a surrogate for changes in the child's behavior that affect exposure to lead in soil, dust, and paint, and also affects the rate of ingestion of these media (Stark et al., 1982; Bornschein et al., 1985; U.S. EPA, 1986, 1989). The amount of soil and dust ingested during normal hand-to-mouth activity tends to a maximum (relative to the size of the child) between ages 15 months and 3 or 4 years for most U.S. children. For risk assessment purposes, the child's age is far more useful in assessing the consequences of changes in exposure. In our reanalyses of the madison County lead data, inclusion of child age as a nonlinear modifying factor improved the goodness of fit of the statistical model significantly. The IEUBK model for lead in children uses age as the basis for adjusting blood lead for the growth of the child, and for changes in ingestion rates, in lead absorption, and in biokinetic parameters. In view of the important differences in lead exposure and uptake for children of different ages that account for differences in blood lead at different ages, even as observed in the IEHR/IDPH report, we believe that further evaluation of age-related differences might have been informative.

**Comment ML5, Section 3.2, p. 3, para. 1-3:** "None of our analyses, besides those involving distance from the smelter, depend in any way upon spatial location.

"Soil lead is not uniformly distributed around the closed smelter either. Although soil lead levels decrease with distance from the closed smelter, there are hot spots and irregularities in the soil lead distribution throughout the sampling area. The sampling areas (z0nes 1--4) were used only to obtain a representative sample of homes and children across the entire range of soil lead levels, regardless of location. Neither distance, nor any other location variable, enters into the main multiple regression/correlation analysis -- the point of which is to use the *joint distribution* of blood, soil, paint, dust and water lead measures in the homes and yards of study participants, regardless of location, to understand how the variables are associated with one another.

"The spatial distribution of blood lead is of interest because it can sometimes help to locate and explain clusters of high blood lead cases."

**Response to Comment ML5:**

These statements appear to be an attempt to justify the investigators' failure to use the extensive collection of previous studies on the site to assist in the design of the Madison County health study. Earlier studies, such as the 1983 Illinois EPA report, clearly identify the non-uniform distribution of soil lead around the NL site, and the existence of isolated areas of higher soil lead in more remote areas of Madison County. These areas often correspond to identifiable sources of contamination from non-smelter sources, and are believed to vary by community. Examples include use of lead-contaminated wastes for street sanding in Venice Township, and use of lead-contaminated waste for yard fill in Eagle Park Acres and elsewhere. Even in areas near the NL site, the distribution of lead in soil was not uniform, as reported by Illinois EPA in 1983, possibly due to differences in patterns of deposition of wind-borne lead particles and of surface water runoff from the site. The investigators appear to have been unaware that lead from different sources may have the potential for different risks to children as a consequence of differences in physical and chemical properties of the lead related to its source. These differences were apparently not considered in their design of the Madison County health study, and we cannot reconstruct this information from the limited data provided to EPA without knowing the location of the sampling sites. The statistical analyses in the IEHR/IDPH report appear to have been carried out as a purely numerical exercise without an understanding of the important physical or chemical properties of environmental media and lead sources that may affect health risk.

**Comment ML6, Sec. 3.2, p. 3, para. 3:**

"The spatial distribution of blood lead is of interest because it can sometimes help to locate and explain clusters of high blood lead cases. That is why we depicted the physical location of the subjects in the study area. However, it was shown that distance is associated not only with soil lead and blood lead, but with SES, building condition, behavior, and other factors that influence blood lead. Simultaneous spatial depiction of all of these factors cannot be interpreted."

**Response to Comment ML6:**

We used distance from the NL site in the EPA reanalyses because it was the only location variable in the data set provided to EPA. However, Table 1 in EPA's comments of May 23, 1994, made extensive use of Figure 1 in the IEHR/IDPH report because it was the only data provided then, or subsequently, that elucidated the differences in spatial location of risk factors for elevated blood lead in children. There is obviously some confounding of distance with building condition and with SES, but much of this confounding could have been accounted for if neighborhood-level information on building condition, paint condition, soil lead, and paint lead had been provided along with the blood lead data. Graphical techniques for displaying multi-dimensional data have been available in many statistical packages for at least a decade, but even simple overlay maps would have been better than providing no information at all about the spatial confounding of potential risk factors.

**Comment ML7, Sec. 3.2, page 3, para. 3:**

"The problem with the unadjusted bivariate tabulation presented by the reviewer in TABLE 1 of the EPA comments is that it totally ignores confounding by these other factors, which we have shown to be present."

**Response to Comment ML7:**

Our bivariate tabulations and graphs in the October, 1994 EPA report were presented to help the reader understand the results. Multivariate statistical methods have been used in our detailed reanalyses of these data, with particular attention to assessment of and adjustments for potential confounding by other variables with spatial gradients across the Madison county study area. We will use appropriate graphics and other multivariate methods in our current reassessment of the data.

**Comment ML8, Sec. 3.3, page 3:** "The reviewer took a meaningful linear multiple regression equation, mistakenly attempted to exponentiate the entire equation, and transformed it into a meaningless expression. The reviewer obviously misunderstood both the use of logs of the environmental and blood lead variables, and the meaning of the original regression equation."

**Response to Comment ML8:**

The reviewers admit that there was a serious typographical error in the equation in nthe report of May 23, 1994; this was corrected in a September 18 draft. We regret any confusion that may have arisen as a result of this. However, the basic point is correct.

The regression equations presented in the IEHR/IDPH report are in general of the form:

log(blood lead) = a + b log(soil lead concentration) + c log(CXRFIAV) + other terms.

CXRFIAV is the average product of interior paint condition and interior lead paint loading measured by a portable X-ray fluorescence analyser (XRF), used in the IEHR/IDPH report without separating the effects of paint condition and lead loading. The estimates of the parameters, here denoted a, b, c, etc., were derived from a least-square fit to the data. We trust that the reviewer will
agree that the logarithm or log function has the property that for two positive numbers, say x and y,
the logarithm of their product is the sum of their logarithms,

$$\log(x\ y) = \log(x) + \log(y),$$

and that if these are natural or base e logarithms whose antilogarithm is the exponential function (denoted by exp), then

$$\exp(\ (\log(x\ y)\ \ ) = \exp\ (\ \log(x) + \log(y)\ ) = \exp(\ \log(x))\ \exp(\ \log(y)\ ) = x\ \ y.$$

We also need the fact that for any number z,

$$z \log(x) = \log(\ x^z).$$

Going back to the original equation,

exp( log(blood lead) ) = exp(a + b log(soil lead concentration) + c log(CXRFIAV) + other t

= exp(a) exp(b log(soil lead concentration)) exp( c log(CXRFIAV) ) exp(other terms)

= exp(a) (soil lead concentration)$^b$ CXRFIAV$^c$ * exp(other terms)

= blood lead.

This is actually a prediction equation for the mean of the exponential (antilogarithm) of log( blood lead), that is, the geometric mean blood lead concentration. The problem arises when this equation is used to estimate the effects of near-perfect remediation in any medium, so that by setting soil lead = 0, the predicted blood lead is also 0 even is paint is not removed. We have, however, also used a linear equation in logarithms as well as a log-transformed linear model for all EPA reanalyses in order to compare different specifications of the prediction equations.